DOCUMENT RESUME

ED 263 170 TM 850 660

AUTHOR	Swarthout, David; And Others
TITLE	The Effect of Geographic Area on General Aptitude Test Battery Validity and Test Scores. USES Test Research Report No. 48.
INSTITUTION	Northern Test Development Field Center, Detroit, MI.
SPONS AGENCY	Employment and Training Administration (DOL), Washington, DC. Div. of Counseling and Test Development.
PUB DATE	84
NOTE	34p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Aptitude Tests; Ethnic Status; *Geographic Regions; Meta Analysis; *Occupational Tests; Racial Differences; Scores; Test Construction; *Test Validity; *Vocational Aptitude
IDENTIFIERS	*General Aptitude Test Battery; Specific Aptitude Test Battery; *United States; United States (North); United States (South); United States (West)

ABSTRACT

Meta-analysis techniques were used to determine whether or not there are differences in General Aptitude Test Battery (GATB) validity and test scores between samples from the Northern, Southern, and Western United States census areas. The sample consisted of 26,111 subjects from 122 validation studies of the Specific Aptitude Test Battery (SATB). Four different methods were used to analyze the differences in validities: (1) comparison of average validities, weighted by sample size; (2) comparison of average differences in validities, weighted by sample size, between geographic areas; (3) comparison of job-family validities; and (4) chi-square (Hunter, Schmidt, and Hunter). The overall conclusion was that there were no meaningful differences in GATB validity between geographic areas. Analysis of mean score differences indicated mean geographic area differences on all nine GATB aptitudes: general learning ability; verbal, numerical, and spatial aptitude; form and clerical perception; motor coordination; and finger and manual dexterity. The Western United States scored highest, with South lowest, and North in between. These differences were not attributed to racial status, since the same pattern was found for Blacks and nonminorities. (Author/GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED263170

USES TEST RESEARCH REPORT NO. 48

THE EFFECT OF GEOGRAPHIC AREA ON GENERAL APTITUDE TEST BATTERY VALIDITY AND TEST SCORES

Analysis and Report
by
Northern Test Development Field Center
Detroit, Michigan

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

DIVISION OF COUNSELING AND TEST DEVELOPMENT
EMPLOYMENT AND TRAINING ADMINISTRATION
U.S. DEPARTMENT OF LABOR
WASHINGTON, D.C. 20213

1984

TM 850 660

TABLE OF CONTENTS

PAGE

LIST OF TABLES i

ACKNOWLEDGMENT ii

ABSTRACT iii

INTRODUCTION 1

SAMPLE 2

PROCEDURE 2

RESULTS 7

DISCUSSION AND CONCLUSIONS 13

REFERENCE NOTES 18

REFERENCES 18

APPENDIX 19

LIST OF TABLES

Table Number	Page	
1a	3	Etnic/minority composition of sample by geographic area.
1b	3	Number of males and females in each geographic area.
2	4	Number of studies and subjects for each criterion type.
3	6	Job family and test battery composition.
4	6	Number of chi-squares significant at the .05 level.
5	8	Cummulative chi-squares.
6	9	Mean differences in validities (weighted by sample size).
7	10	Average validities (weighted by sample size).
8	10	Critical ratios between mean weighted validities.
9	11	Mean weighted validities of job family norms for studies in each job family by geographic area.
10	12	Critical ratios between mean weighted validities for each geographic area for each job family.
11	14	Analysis of variance results for total sample.
12	15	Analysis of variance results for blacks.
13	16	Analysis of variance results for nonminorities.

ACKNOWLEDGMENT

The United States Employment Service (USES) conducts a test research program for developing testing tools useful in vocational counseling and placement.

The purpose of this series of reports is to provide results of significant test research projects as they are completed. These reports will be of interest to users of USES tests and to test research personnel in state agencies and other organizations.

The computer analysis was conducted and the report written by David Swarthout, David Synk, and William Goode of the Northern Test Development Field Center, Detroit, Michigan.

ABSTRACT

This study used meta-analysis research techniques to determine if there are differences in General Aptitude Test Battery (GATB) validities and test scores between samples from the Northern, Southern, and Western census areas. The sample consisted of 26,111 subjects from 122 Specific Aptitude Test Battery (SATB) validation or revalidation studies analyzed since 1972.

Four approaches were used to analyze the differences in validities between geographic areas. The first approach compared average validities, weighted by sample size. The largest difference found was .03 correlation points. The second approach compared average differences in validities, weighted by sample size, between geographic areas. The average difference in validities for all aptitudes across all area comparisons was .008 correlation points. The third approach compared job family validities (Hunter, Note 1) and showed no significant differences. The fourth approach used the chi-square from Hunter, Schmidt, and Hunter (1979). This analysis showed slightly more significant differences than would be expected by chance, but this could be due to three sources of Type I error bias described by Hunter, et al. (1979) that were present in this study.

The overall conclusion of this study is that there are no meaningful differences in GATB validities between geographic areas.

Mean score differences were also analyzed. There were mean differences between the geographic areas on all nine GATB aptitudes, with the West highest, South lowest and North in between. These differences cannot be attributed solely to the differences in racial composition between areas, since the same pattern of differences was found for blacks and nonminorities.

The results have implications for the present geographic sampling requirements for SATB research and the transportability of validity.

INTRODUCTION

The U.S. Employment Service (USES), in cooperation with State employment services, has conducted a continuing program of occupational test research and development since the mid 1930's. Most of this effort has been devoted to developing and researching the General Aptitude Test Battery (GATB). The GATB consists of 12 tests measuring the following nine vocationally-relevant aptitudes:

General Learning Ability (G)
Verbal Aptitude (V)
Numerical Aptitude (N)
Spatial Aptitude (S)
Form Perception (P)
Clerical Perception (Q)
Motor Coordination (K)
Finger Dexterity (F)
Manual Dexterity (M)

The validation of the GATB for specific occupations has resulted in the development of over 470 Specific Aptitude Test Batteries (SATBs). These batteries consist of combinations of two, three, or four GATB aptitudes with associated cutting scores. All of the SATBs were developed from empirical research studies. In each study criterion data measuring job proficiency were collected along with GATB test scores. The validity of the aptitudes was measured by the correlation between aptitude test scores and the criterion.

One issue that the USES has been concerned with is what variables affect or moderate GATB validities. Some of the variables that have been postulated to moderate test validity are minority group status, sex, geographic area, age, education and work experience. The present study looks at one of these variables - geographic area. The study uses meta-analysis research techniques on SATB validation data to determine if there are differences in validities and test scores between samples from the Northern, Southern and Western census areas.

In the early 1970's the USES consulted with the Equal Employment Opportunity Commission (EEOC) regarding compliance of USES tests with the EEOC Guidelines on Employee Selection Procedures (U.S. EEOC, 1970). EEOC representatives questioned the comparability of USES samples in the North with those in the South and West. As a result of those meetings, the USES embarked on a program to revalidate its SATBs with separate analyses on minority and nonminority groups as well as the three census areas (North, South and West).

The issue of whether geographic area moderates test validity has important implications for the conduct of GATB validity research and is of interest to both psychologists who develop tests and employers who are concerned with test costs and benefits. Psychologists are interested in the variable of geography for technical reasons: (1) The Uniform Guidelines on Employee Selection Procedures (U.S. EEOC, 1978) indicate that test transportability is permitted when certain conditions are met. In the case of criterion-related validity, samples should be comparable in terms of "relevant factors likely to affect validity differences...." (p. 38301). Geography is one factor thought to affect validity. (2) Schmidt and Hunter (1977) have found that for a variety of test types and jobs, most or all observed evidence for situational specificity is artifactual in nature. It follows that geography probably has a minimal effect on validity.

In addition to technical issues, the concept of geography moderating validity has practical implications: (1) It is important to all large multistate employers and those organizations involved in cooperative validity research because requiring samples with geographic representation greatly increases the costs of validity studies. (2) It is of interest to both large and small employers who look to the Job Service for test-selected applicants because it affects the confidence they can place in GATB validity studies for development of SATBs done in geographical areas other than those in which applicants are tested for job orders.

Only one citation pertaining to geographic differences was found from computerized and non-computerized literature searches (Department of Defense, Note 3). This study found differences in mean aptitude scores between geographic areas for the Armed Forces Qualification Test (AFQT), with youth in the North scoring highest and youth in the South scoring lowest.

SAMPLE

The sample consists of available data from SATB validation or revalidation studies analyzed since 1972. The total N is 26,111 from 122 studies. One hundred and fourteen studies (N=23,558) used a concurrent validation design and eight studies (N=2,553) used a longitudinal design. Table 1 gives minority and sex breakdown by geographic area.

Most of the criterion data consisted of the sum of scores from two administrations of the Standard Descriptive Rating Scale. The scale was used to get ratings from supervisors on five aspects of job performance (quantity, quality, accuracy, job knowledge and job versatility) as well as "all-around" ability. However, other types of criterion data were collected and criteria were combined in different ways (see Table 2). The Appendix contains a listing of all 122 studies and the number of individuals in each geographic area.

PROCEDURE

In each study, validities for the nine GATB aptitudes were computed for each geographic area that had more than 25 subjects. Differences in validities were analyzed for each area comparison (North-South, North-West, and South-West).

TABLE 1A

Ethnic/Minority Composition of Sample by Geographic Area

N=25,288[a]

	North N=10,111	South N=10,116	West N=5,061
Black	2,648	4,002	623
American Indian	165	77	166
Asian	131	54	224
Hispanic	348	381	992
White	6,819	5,602	3,056

[a]Information unavailable for 823.

TABLE 1B

Number of Males and Females in each Geographic Area

N=26,111

	North N=10,177	South N=10,782	West N=5,152
Male	4,814	5,071	2,531
Female	5,363	5,711	2,621

TABLE 2

Number of Studies and Subjects for Each Criterion Type

Type of Study	# of Studies	# of Individuals	Criterion Measure		
			CR1	CR2	CR3
Standard DRS Concurrent	94	19,942	Standard DRS	Standard DRS	Sum of CR1 and CR2
Two criteria collected but only one used as final	1	119	Special DRS	Work Sample	CR2
Multiple Hurdle	2	647	Standard DRS	Course grades	
	1	933	Special DRS	Standard DRS	
	1	95	Course grades	Standard DRS	
Final criterion is combination of different criteria	1	213	Special DRS	Ranking	Combination of CR1, CR2
	1	81	Standard DRS	Performance Model	Combination of CR1, CR2
	1	123	Special DRS	Standard DRS	Combination of CR1, CR2
	1	286	Mixed Standard	Standard DRS	Combination of CR1, CR2
Final criterion is combination of same criteria (not Standard DRS)	1	119	Course grades	Course grades	Sum of CR1 and CR2
	3	723	Broad category rating	Broad category rating	Sum of CR1 and CR2
	1	141	Ranking	Ranking	Sum of CR1 and CR2
	10	1,902	Special DRS	Special DRS	Sum of CR1 and CR2
Longitudinal Standard DRS	1	81	Standard DRS	Standard DRS	Sum of CR1 and CR2
Only one criterion	1	270	————	Special DRS	————
	1	107	Mixed Standard Rating Scale	————	————
	1	329	Course grades	————	————
	122	26,111			

Validities were analyzed in two ways for the studies that had more than 25 subjects from each of two or three geographic areas. There were 74 of these studies for North-South; 51 for North-West; and 49 for South-West. The first method for analyzing validities for these studies is based on the Hunter, Schmidt and Hunter (1979) analysis of differential validity. This method involves computing the following chi-square for each validity pair.

$$\chi^2 = \frac{(Z_1 - Z_2)^2}{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}$$

where Z_1 and Z_2 are the validity coefficients in Fisher Z form for area 1 and area 2, and N_1 and N_2 are sample sizes for area 1 and 2. The number of significant chi-squares was computed as well as the cumulative chi-square.

Also, the difference in validities between geographic areas was computed for each of these studies and weighted by the combined sample size for the two areas. The mean of these weighted differences was computed for the nine aptitudes for each area comparison.

There were 37 studies that had more than 25 subjects from just one geographic area. These studies were combined with the studies that had more than 25 subjects from each of two or three areas. Validities for this larger data set were analyzed in two ways. The first compared validities weighted by sample size. Average validities, weighted by sample size, were computed for each geographic area and tested for differences. The second analysis compared job family validities (Hunter, Note 1) between geographic areas. Hunter developed a grouping system of jobs based on the Data and Things ratings of the occupational codes in the Dictionary of Occupational Titles (DOT) (United States Department of Labor, 1977). Each job in the DOT is in one of the five job families (see Table 3). The regression weights for each job family were used to get predicted criterion scores and the correlation was computed between predicted and actual criterion scores for each study. Average validities, weighted by sample size, for each of the five job families were computed for each geographic area and critical ratios computed between the areas.

Mean scores were compared by computing simple one way analyses of variance (ANOVA) on the total sample between geographic areas for sex, age, education, plant experience, total experience and the nine GATB aptitudes. ANOVAs were also computed separately for black and nonminority subgroups.

TABLE 3

Job Family and Test Battery Composition

Job Family	Contribution to Composite: Cognitive GVN	Perceptual SPQ	Psychomotor KFM	DOT Data-Things Code
1	59%	30%	11%	T=0=Setting up
2	13%		87%	T=6=Feeding-Offbearing
3	100%			D=0=Synthesizing =1=Coordinating
4	73%		27%	D=2=Analyzing =3=Compiling =4=Computing
5	44%		56%	D=5=Copying =6=Comparing

TABLE 4

Number of Chi Squares Significant at .05 Level

	Number of Studies	G	V	N	S	P	Q	K	F	M	Total
North-South	74	5	3	6	6	11	8	8	13	9	69
North-West	51	2	2	3	4	4	7	4	4	5	35
South-West	49	2	3	2	5	5	5	4	1	0	27
Total	174	9	8	11	15	20	20	16	18	14	131

RESULTS

Table 4 shows the number of chi-squares significant at the .05 level for each area comparison. When a number of significance tests are performed, a certain number would be expected to be significant by chance. That is, if a number of samples were drawn from a population for which it was known that there are no differences between groups, for some of these samples the tests of differences would be significant. Determining how many tests of differences would be expected to be significant if there are no differences in the population is discussed by Brozek and Tiede (1952) and Cross and Chaffin (1982). The binomial probability distribution was used to determine the number of tests expected to be significant when up to 49 studies are involved. For 50 or more studies the normal distribution was used to determine the number of tests expected to be significant. The number of chi-squares expected to be significant for the 74 North-South chi-squares is 7.4; for the 51 North-West chi-squares it is 5.6; and for the 49 South-West chi-squares it is 5.4. Only aptitudes P, Q, K, F and M for North-South and aptitude Q for North-West have more chi-squares significant than would be expected by chance.

Table 5 shows the cumulative chi-squares for each aptitude for each area comparison and row and column sums. The cumlative chi-squares for each area comparison and total cumulative chi-square are significant. The results for the total chi-square should be viewed with caution because the samples are not independent.

Table 6 shows the weighted mean differences in validities for each aptitude for each area comparison. The weighted mean difference for all aptitudes for all comparisons is -.008. These comparisons between areas were made in this order - North-South, North-West and South-West. A minus sign indicates that the coefficient for the second area in the pair was larger. For example, the mean difference for G between North-South was - .014, which means that on the average, the validity coefficient for the South was .014 higher than the North.

Average validities, weighted by sample size, for each area are shown in Table 7. These validities include data from studies with more than 25 subjects from only one geographic area. Inspection of Table 7 shows that for each aptitude the validities for the three areas are virtually the same. The critical ratios between these validities are shown in Table 8. Only aptitude S between North and South showed a significant difference and that difference was only .03 correlation points.

The results for the fourth method of analyzing differences in validities based on Hunter (Note 1) are shown in Tables 9-10. Each study was placed in one of the five job families based on Data and Things ratings of their DOT codes. Job families 2 and 3 had only one study each and all the data were collected from one geographic area. The critical ratios between areas are shown in Table 10. None of the critical ratios were significant.

TABLE 5

Cumulative Chi-Squares

	Number of Studies	G	V	N	S	P	Q	K	F	M	Total
North-South	74	86.14	82.0	93.89	86.99	119.27**	113.38**	109.12**	140.98**	114.66**	946.43**
North-West	51	56.26	57.17	55.21	54.68	69.88*	77.63**	89.70**	71.18*	66.36	598.07**
South-West	49	53.08	63.50	50.40	57.64	62.06	66.68*	55.11	45.15	42.97	496.59*
Total	174	195.48	202.67	199.50	199.31	251.21**	257.69**	253.93**	257.31**	223.99**	2041.09**

*Significant at .05 level.
**Significant at .01 level.

TABLE 6

Mean Differences in Validities (Weighted by Sample Size)

	North-South N=14,018	North-West N=9,933	South-West N=7,859
G	-.014	-.008	.005
V	-.011	-.007	.009
N	.002	-.010	-.021
S	-.013	-.005	.015
P	.010	-.024	-.028
Q	.016	.003	-.039
K	-.014	-.047	-.008
F	.015	-.014	-.017
Z	.012	-.038	.007

TABLE 7

Average Validities
(Weighted by Sample Size)

	North N=9,963	South N=10,657	West N=4,804
G	.21	.23	.22
V	.16	.18	.16
N	.22	.21	.22
S	.13	.16	.14
P	.16	.15	.17
Q	.18	.16	.16
K	.11	.10	.11
F	.12	.10	.10
M	.11	.11	.12

TABLE 8

Critical Ratios Between Mean Weighted Validities

	G	V	N	S	P	Q	K	F	M
North-South	-1.51	-1.00	.36	-2.22*	1.08	1.22	.86	1.00	.14
North-West	-.34	.06	-.34	-.28	-.51	1.25	-.11	1.14	-.51
South-West	.86	.86	-.63	1.50	-1.38	.29	-.81	.35	-.63

*Significant at .05 level.

TABLE 9

Mean Weighted Validities of Job Family Norms

for Studies in Each Job Family by Geographic Area

	Total			North			South			West		
Job Family	Number of Studies	N	Validity	Number of Studies	N	Validity	Number of Studies	N	Validity	Number of Studies	N	Validity
1	5	906	.15	3	244	.17	5	415	.18	3	201	.07
2	1	126	.09	—			1	126	.09	—		
3	1	75	.28	—			1	75	.28	—		
4	67	16179	.24	55	6778	.23	53	5529	.26	37	3520	.24
5	48	8825	.21	30	2941	.23	44	4512	.20	19	1083	.22

TABLE 10

Critical Ratios Between Mean Weighted Validities
For Each Geographic Area For Each Job Family

Job Family	North-South	North-West	South-West
1	-.18	.98	1.25
2	---	---	---
3	---	---	---
4	-1.30	-.49	.62
5	1.25	.15	-.71

The results for the comparison of mean scores are shown in Table 11. There were significant differences between areas for all variables. The same pattern of differences was found for all nine GATB aptitudes - West high, South low, and North in between. The same pattern of differences was found for blacks and nonminorities (see Tables 12-13).

DISCUSSION AND CONCLUSIONS

The primary purpose of this study was to find out if the validity of the GATB is the same in the Northern, Southern and Western areas of the United States. Four different types of validity analyses were performed. Overall the data strongly support the conclusion that there is no difference in validity between different areas of the country.

The two most relevant analyses are those presented in Tables 7-8 and 6. The data in Tables 7-8 compared SATB average validities, weighted by sample size, across all three areas. This analysis used 97% of the total sample while the other two analyses used less than this. The interpretation of these results is clear. There are no meaningful differences in area validities. The data in Table 6 compared differences in validities on a study-by-study basis, with the results averaged over all comparisons. The average difference for all aptitudes across all area comparisons is -.008.

The above two analyses are the most relevant for another reason. Both analyses report results in terms of correlation units. This allows for the comparison of the magnitude of differences in validities. Some of the other analyses report only that the differences were greater than expected by chance and nothing about the magnitude of the differences in terms of validity coefficients.

Even though the other two validity analyses are not quite as pertinent, they still support (though less strongly) the hypothesis of no geographic differences. One explanation for the slightly weaker support provided by these two analyses is the existence of three sources of Type I error bias that make the .05 level of significance inappropriate. The three sources of Type I error bias and their effect on the interpretation of the significance level were discussed by Hunter et al. (1979).

The first source of Type I error bias is the effect of nonnormal test score distribution due to selecting applicants using the same test, or a similar test, to that used for the validation study. To the extent that this happens, the result is to increase the probability of falsely finding differences in validities beyond the specified .05 level. This particular source of Type I error bias did not affect the results of this study very much since most of the samples with workers test selected on instruments similar to the GATB were eliminated.

The second source of Type I error bias results from the lack of complete independence between validity pairs. Predictor and/or criterion redundancy increases Type I error, although it is difficult to calculate the effect. This source of bias occurs in the present study because the nine GATB aptitude scores (which are correlated with each other) are correlated with the same criterion to obtain the nine validity coefficients. Since the present analyses were composed of three separate comparisons (North-South, North-West, and South-West), each sample was included <u>twice</u>.

TABLE 11

Analysis of Variance Results for Total Sample

	North N=10,177[a]		South N=10,782[b]		West N=5,152[c]		
	Mean	SD	Mean	SD	Mean	SD	F
Sex	1.5	.5	1.5	.5	1.5	.5	3.27*
Age	32.5	11.6	30.5	10.0	32.3	11.1	94.81**
Ed	12.2	1.7	11.8	1.8	12.4	1.8	283.38**
Plt	63.1	73.9	44.4	52.3	46.5	57.9	226.86**
Tot	82.0	86.7	54.6	63.9	76.0	86.8	325.33**
G	97.4	18.4	90.8	18.5	99.7	19.2	531.02**
V	98.3	15.9	91.7	15.0	99.7	16.4	660.90**
N	96.5	19.6	90.3	19.9	96.6	19.8	317.78**
S	99.8	19.5	95.9	19.7	104.7	20.0	355.23**
P	106.9	22.2	102.1	21.7	110.9	22.5	306.37**
Q	112.6	18.7	106.7	18.0	113.8	18.5	385.84**
K	105.4	19.3	102.6	18.8	105.9	19.2	79.13**
F	97.8	22.1	96.5	21.1	99.2	22.4	29.96**
M	105.3	23.0	103.7	21.1	107.7	23.5	56.14**

[a]N for plant experience is 8786, N for total experience is 8908.
[b]N for plant experience is 9469, N for total experience is 10484.
[c]N for plant experience is 4699, N for total experience is 5026.

*Significant at .05 level.
**Significant at .01 level.

TABLE 12

Analysis of Variance Results for Blacks

	North N=2,648[a]		South N=4,002[b]		West N=623[c]		
	Mean	SD	Mean	SD	Mean	SD	F
Sex	1.6	.5	1.6	.5	1.7	.5	4.24*
Age	32.5	10.2	29.3	8.5	31.8	9.4	101.85**
Ed	12.1	1.5	11.9	1.6	12.5	1.7	36.44**
Plt	56.7	59.5	39.3	41.7	45.4	49.7	90.72**
Tot	73.0	70.1	47.0	50.2	67.6	61.8	153.46**
G	84.4	15.6	79.9	15.1	84.5	16.3	75.21**
V	89.1	13.2	84.5	11.7	88.7	12.9	120.59**
N	84.3	18.5	79.9	18.3	83.6	18.6	47.00**
S	89.6	17.5	86.9	17.0	90.5	17.9	25.66**
P	96.7	22.3	95.5	21.2	99.5	22.3	9.93**
Q	105.0	18.8	102.4	17.0	105.8	18.1	22.86**
K	102.9	20.2	102.2	19.2	102.7	18.7	1.05
F	91.6	21.7	93.1	21.0	92.3	21.0	4.01*
M	99.6	22.0	102.2	20.5	101.4	21.8	11.87**

[a]N for plant experience is 2424, N for total experience is 2440.
[b]N for plant experience is 3750, N for total experience is 3874.
[c]N for plant experience is 611, N for total experience is 619.

*Significant at .05 level.
**Significant at .01 level.

TABLE 13

Analysis of Variance Results for Nonminorities

	North N=6,819[a]		South N=5,602[b]		West N=3,056[c]		
	Mean	SD	Mean	SD	Mean	SD	F
Sex	1.5	.5	1.5	.5	1.5	.5	.48
Age	32.6	12.2	31.8	10.9	32.4	11.5	8.39**
Ed	12.3	1.8	11.7	1.9	12.6	1.7	274.12**
Plt	67.4	80.2	48.0	58.4	45.0	59.3	151.27**
Tot	87.3	93.6	61.7	71.7	80.9	94.3	130.51**
G	102.8	16.6	97.4	17.4	106.0	16.8	294.19**
V	102.4	15.1	96.6	15.1	104.9	15.3	369.33**
N	101.6	17.7	96.8	18.3	102.1	17.5	138.49**
S	103.9	18.7	100.8	19.1	109.2	18.8	196.56**
P	111.1	21.0	106.9	21.2	115.4	20.9	168.67**
Q	116.0	17.7	111.5	17.3	117.3	17.7	145.43**
K	106.2	18.9	103.4	18.1	106.6	18.8	43.66**
F	99.8	22.0	98.1	21.3	100.0	22.4	11.86**
M	107.1	23.2	104.3	21.7	108.5	23.7	39.79**

[a]N for plant experience is 5824, N for total experience is 5883.
[b]N for plant experience is 5218, N for total experience is 5439.
[c]N for plant experience is 2781, N for total experience is 2980.

*Significant at .05 level.
**Significant at .01 level.

The third source of Type I error bias is due to differential range restriction on the predictor variables. This has the effect of showing a statistically significant difference between two scores that would not be statistically significant if the extent of range restriction had been equal on both variables. Thus, in some cases, this results in falsely rejecting the null hypothesis and so increases the "real" alpha level to some value above the assigned .05 level. There is some differential range restriction of predictor scores across the different geographic regions but in this study the overall increase in Type I error due to this source of bias is probably not great.

Because of the very large sample size used in this study, (N=26,111) the probability of detecting true differences when they really exist (statistical power) is extremely high. An examination of the results of this study shows that most of the statistically significant differences between areas were small in magnitude and were detected due to high statistical power associated with this sample.

The impact of having high statistical power and increased Type I error bias is as follows: When we failed to detect differences in validity between areas, we are quite sure that such differences really do not exist because of the high power of our test. But, when we find significant differences between areas, we cannot be sure if they are real differences, or were detected due to the Type I error biases. Also, even if we were sure that such differences were true statistically significant differences, they are too small to be meaningful and practically useful.

It is clear that there are differences between geographic areas in mean aptitude scores. These differences cannot be attributed solely to the differences in racial composition between areas, since the same pattern of differences was found for blacks and nonminorities. One explanation may be that areas differ in factors such as urban-rural composition, quality of education and socio-economic status that Tyler (1965) found to affect aptitude scores.

The theoretical implications of this study support the contention of Schmidt and Hunter (1977) that most observed variance of validities is artifactual. That is, validities are more stable than previously thought, and are transportable across such diverse variables as job grouping, race, and geographic regions.

The practical implications of this study for the USES testing program are twofold. First, since there are no meaningful differences in validities between different geographic areas, the requirements for regional sampling for SATB Test Development Projects should be eliminated. This would reduce the time and effort involved in developing SATBs. Second, since there is no evidence of differences between areas on job family validities, the concern over geography as a moderator of validity, would not be a barrier to expanding the North Carolina Validity Generalization Project to other parts of the country.

REFERENCE NOTES

1. Hunter, J.E. Test validation for 12000 jobs: An application of job classificaton and validity generalization analysis to the General Aptitude Test Battery (GATB). U.S. Employment Service, 1980.

2. U.S. Department of Defense, Profile of american youth, 1982.

REFERENCES

Brozek, J. & Tiede, K. Reliable and questionable significance in a series of statistical tests. Psychological Bulletin, 1952, 49, 339-341.

Cross, E.M. & Chaffin, W.W. Use of the binomial theorem in interpreting results of multiple tests of significance. Educational and Psychological Measurement, 1982, 42, 25-34.

Hunter, J.E., Schmidt, F.L., & Hunter, R. Differential validity of employment tests by race: A comprehensive review and analysis. Psychological Bulletin, 1979, 86, 721-735.

Schmidt, F.L., & Hunter, J.E. Development of a general solution to the problem of validity generalization. Journal of Applied Psychology, 1977, 62, 529-540.

Tyler, L.E. The psychology of human differences (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall, Inc., 1965.

U.S. Department of Labor, Dictionary of occupational titles (4th ed.). Washington, D.C., 1977.

U.S. Equal Employment Opportunity Commission. Guidelines on employee selection procedures. Federal Register, 1970, 35 (149), 12333-12336.

U.S. Equal Employment Opportunity Commission, U.S. Civil Service Commission, U.S. Department of Labor, & U.S. Department of Justice. Uniform guidelines on employee selection procedures. Federal Register, 1978, 43 (166), 38295-38309.

APPENDIX

DOT Title and Code and Number of Subjects from Each Geographic Area

SATB No. or Study No.	DOT Title	DOT Code	Number of Subjects			
			Total	North	South	West
2	Stock Clerk	222.387-058	151	55	77	19
4	Sewing Machine Operator	787.682-046	208	25	132	51
7	Laboratory Tester	029.261-010	95	53	33	9
9	Central Office Operator	235.462-010	102	13	79	10
10	Stenographer	202.362-014	622	415	76	131
10	Clerk-Typist	203.362-010	431	226	87	118
10	Typist	203.582-066	141	108	22	11
11	Carpenter	860.381-022	154	105	16	33
11	Carpenter	860.381-022	119	46	73	0
12	Machinist	600.280-022	283	154	45	84
28	Packager, Hand	920.587-018	445	376	33	36
31	Checker II	209.687-010	121	80	0	41
31	Checker II	209.687-010	59	59	0	0
34	Bindery Worker	653.685-010	185	133	48	4
38	File Clerk II	206.367-014	211	85	126	0
43	Automobile Mechanic	620.261-010	425	144	148	133
44	Punch-Press Operator I	615.482-022	89	34	42	13
45	Shipfitter	806.381-046	252	82	122	48
47	Nursery School Attendant	359.677-018	174	57	66	51

SATB No. or Study No.	DOT Title	DOT Code	Number of Subjects			
			Total	North	South	West
53	Spinner, Frame	682.685-010	180	11	169	0
61	Plumber	862.381-030	253	100	118	35
61	Plumber	862.381-030	411	0	411	0
63	Garment Folder	789.687-066	103	10	93	0
68	Refinery Operator	549.260-010	194	63	90	41
71	Cosmetologist	332.271-010	386	0	0	386
72	Electrician	824.261-010	394	50	161	183
72	Electrician	824.261-010	253	0	253	0
74	Central-Office Repairer	822.281-014	142	30	103	9
80	Radiologic Technologist	078.362-026	137	14	41	82
82	Sheet-Metal Worker	804.281-010	152	91	61	0
101	Assembler, Automobile	806.684-010	213	119	94	0
115	Weaver	683.682-038	126	19	98	9
120	Fire Fighter	373.364-010	130	0	130	0
124	Tractor-Trailer-Truck Driver	904.383-010	320	97	114	109
126	Welder, Combination	819.384-010	220	56	38	126
131	Industrial-Truck Operator	921.683-010	202	107	69	26
135	Production-Machine Tender	609.685-010	227	96	113	18
141	Bench Assembler	706.684-642	160	18	142	0
144	Machinist, Wood	669.380-014	100	27	52	21
145	Cashier-Checker	211.462-014	119	0	28	91

SATB No. or Study No.	DOT Title	DOT Code	Number of Subjects			
			Total	North	South	West
153	Loom Fixer	683.260-018	156	16	140	0
165	Packager, Hand	920.587-018	102	32	70	0
168	Yarn-Texturing-Machine Operator	589.685-102	111	21	90	0
177	Millwright	638.281-018	302	165	113	24
180	Keypunch Operator	203.582-030	353	194	131	28
182	Laborer, Stores	922.687-058	127	30	64	33
199	Audit Clerk	210.382-010	300	218	18	64
200	Reservations Agent	238.367-018	310	213	40	57
207	Welder, Arc	810.384-014	162	27	126	9
211	Welder, Arc	810.384-014	461	305	118	38
211	Welder, Arc	810.384-014	81	81	0	0
217	Proof-Machine Operator	217.382-010	243	43	89	111
220	Coil Winder	724.684-026	115	64	51	0
228	Injection-Molding-Machine Tender	556.685-038	155	38	70	47
231	Surgical Technician	079.374-022	250	134	102	14
235	Metal Fabricator	619.360-014	173	9	88	76
236	Police Officer I	375.263-014	121	0	121	0
238	Cook	313.361-014	114	34	62	18
239	Ward Clerk	245.362-014	185	52	74	59
259	Teller	211.362-018	291	106	77	108

SATB No. or Study No.	DOT Title	DOT Code	Number of Subjects			
			Total	North	South	West
266	Drafter, Civil Drafter, Geological Drafter, Mechanical Drafter, Structural	005.281-010 010.281-018 007.281-010 005.281-014	326	39	36	251
267	Tire Builder, Automobile	750.384-010	239	95	53	91
270	Nurse, Licensed Practical	079.374-014	204	81	108	15
274	Food-Service Worker, Hospital	355.677-010	170	41	81	48
276	Salesperson, General Merchandise	279.357-054	171	89	82	0
278	Sales Clerk	290.477-014	163	48	72	43
280	Structural-Steel Worker	801.361-014	249	185	10	54
281	Electronics Assembler	726.684-018	56	24	32	0
282	Nurse Aide	355.674-014	136	0	136	0
286	Computer Operator	213.362-010	213	91	38	84
287	Psychiatric Aide	355.377-014	334	182	81	71
309	Proof-Machine Operator	217.382-010	172	94	78	0
310	Electronics Assembler	726.684-010	185	46	71	68
313	Automobile-Body Repairer	807.381-010	107	47	3	57
326	Respiratory Therapist	079.361-010	496	281	76	139
327	Psychiatric Technician	079.367-022	384	101	154	129
329	Administrative Clerk	219.362-010	407	135	172	100
330	Chemical Operator III	559.382-018	62	56	0	6

SATB No. or Study No.	DOT Title	DOT Code	Number of Subjects			
			Total	North	South	West
332	Hotel Clerk	238.362-010	406	228	101	77
335	Extruding-Machine Operator	691.382-010	142	26	79	37
336	Knitting-Machine Operator	685.665-014	209	86	123	0
342	Water-Treatment-Plant Operator	954.382-014	222	147	38	37
343	Operating Engineer	859.683-010	90	0	0	90
343	Operating Engineer	859.683-010	270	128	40	102
360	Yarn Winder	681.685-154	207	0	207	0
363	Maintenance Repairer, Factory or Mill	899.281-014	233	85	89	59
370	Maintenance Mechanic	638.281-014	141	117	10	14
375	Lather	842.361-010	114	17	25	72
376	Mailing-Machine Operator	208.462-010	128	45	78	5
379	Transportation Agent	912.367-014	131	120	11	0
381	Electronics Assembler	726.684-018	100	7	89	4
393	Hospital-Admitting Clerk	205.362-018	178	78	78	22
398	Teacher Aide II	249.367-074	266	95	153	18
407	Quality Control Technician	529.387-030	152	41	10	101
414	Assembler, Electrical Accessories I	729.687-010	191	116	52	23
427	Spooler Operator, Automatic	681.686-018	126	0	126	0
436	Food-Service Worker, Hospital	355.677-010	127	58	26	43
447	Welder, Production Line	819.684-010	177	119	49	9

SATB No. or Study No.	DOT Title	DOT Code	Number of Subjects			
			Total	North	South	West
456	Assembler, Small Products	739.687-026	183	50	114	19
465	Covering-Machine Operator	681.685-038	65	0	65	0
466	Material Handler	929.687-030	44	0	44	0
467	Electronics Assembler	726.684-018	276	76	153	47
468	Cigarette Inspector	529.567-010	64	0	64	0
469	Chemical Operator II	558.585-014	246	84	149	13
470	Weaver	769.684-050	81	0	81	0
471	Electronics Inspector	726.684-022	644	0	644	0
472	Appliance Assembler, Line	827.684-010	107	0	107	0
473	Gambling Dealer	343.467-018	933	933	0	0
473	Gambling Dealer	343.467-018	123	0	0	123
1001	Central-Supply Worker	381.687-010	431	218	128	85
1002	Data Typist	203.582-022	174	83	34	57
1003	Etched-Circuit Processor	590.684-018	258	105	121	32
1004	Cytotechnologist	078.281-010	131	65	39	27
1005	Assembler	723.684-010	91	10	81	0
1006	Machine Operator II	619.685-062	247	132	103	12
1007	Supervisor	529.137-026	75	0	75	0
1008	Power-Reactor Operator	952.362-022	329	0	329	0
1009	Customer-Service Representative	239.367-010	278	132	90	56

SATB No. or Study No.	DOT Title	DOT Code	Number of Subjects			
			Total	North	South	West
1010	Meter Reader	209.567-010	286	65	145	76
1011	Packager, Hand	920.587-018	203	41	40	122
1012	Environmental-Control-System-Installer-Servicer	637.261-014	262	0	262	0
3048	Pipe Fitter	862.261-010	95	95	0	0